Snežana Popović
The Faculty of Computer Science, Belgrade, Serbia
Nina Krstić
Museum of Naïve and Marginal Art (MNMA), Jagodina, Serbia
Zoran Cvetković
Infotrend, Belgrade, Serbia

EFFECTIVE PUBLISHING OF DIGITALIZED CULTURAL HERITAGE ON THE WEB

Abstract.T he process of digitization of cultural property has multiple targets, defined by the strategy of cultural institutions. First, digitization contributes to the protection of cultural properties by obtaining all related data and their representative archive images. Besides, images of digitalized properties have a significant role in the online presentation and promotion of cultural artefacts. This paper presents our research and cognition about prerequisites and promising ways of successful presentation and promotion of digitized cultural artefacts on the Web. The fact is that today's Web is an immeasurable repository of data and images. Thus, it is difficult to achieve satisfactory promotional results, if images of cultural artefacts are not described on the Web.In addition, the benefits of available technologies and vocabularies, applicable for effective semantic annotation of images were studied. We concluded that the use of Schema.org and Getty vocabulary, followed by JSON-LD serialization, contributed to more accurate description and categorization of cultural property images. We applied and checked the conclusions of our research within the project "Marginal Art in Serbia". The project examines professional interpretation and Web promotion of marginal artists' artworks from the collection of Museum of Naïve and Marginal Art (MNMA), Jagodina. The review of the Project's final results confirms that Google recognized our semantically annotated images as structured data, classifying them into the related categories and assigning them high ranking in its indexes.

Keywords: schema.org, Linked Open Data, semantic enrichment, effective annotation, artwork images

1 Introduction

The status of "cultural heritage" arises from cultural significance and values of an artefact for closer or wider environment [1]. In this article we present conclusions of our research about the preconditions for successful presentation and promotion of cultural artefacts on the World Wide Web. From our point of view, the "successful promotion" means understanding and acceptance of the values and importance of cultural property, based on expert interpretation and presentation, as well as acceptable rank in the Google search engine indexes.

The benefits of Web environment for promotion of cultural properties can be functional if the digitized cultural properties are semantically annotated in an adequate manner. We decomposed the process of the "adequate" semantic annotation of digitized cultural properties into two groups of activities: description and technical implementation.

In addition to the curator's expertise and creativity, the description, or semantic annotation of a cultural property includes the use of thematic vocabularies. Within our study we examined available domain vocabularies on the Web [2]. Based on our research, we prefer to use Getty and Schema.org vocabularies.

Technical implementation of description of digitized cultural property requires the use of applicable standards and technologies for semantic annotation of digital contents on the Web. Within our research, we found that there were multiple parallel standards and applicable technologies which are discussed in the Chapter 3 of this article.

The choice of vocabulary, technology and standards for semantic annotation of digitized cultural artefacts is not unique. Therefore, we made efforts to discover a combination that described cultural heritage adequately and also had positive impact on the ranking of digitized cultural properties in search engine indexes.

Our conclusions and recommendations were implemented in the Project "Marginal Art in Serbia". The obtainedresults confirm that the proposed LOD online vocabularies together with the applied standards and technologies enable adequate interpretation of layered messages of marginal artworks and get high positioning in the indexes of the Google search engine.

2 Interpretation and promotion of artworks on the web

Available technical devices, computer equipment and software make it easy to photograph and scan cultural heritage, thus improving their digitization [3]. Also, simple publishing methods made Web ahuge repository of textual and visual contents, including digitized cultural properties as well.However, visual contents of images of cultural artefacts published on the Web are not always enoughfor their proper interpretation and understanding. The lack of understanding is particularly noticeable when presenting images of artworks from some categories of fine art.

To present and promote a digitized cultural artefact, even in the competitive Web environment, it is necessary to annotate its image in a way that is understandable to people and machines.One of the recommended frameworks for describing cultural properties includes the analysis of visual impact, emotional expression and its symbolic content [4]. A good way to implement the recommended framework is the use of the accepted terms in the form of metadata. On the Internet, there are already available structured vocabularies. They direct the semantic annotation of cultural properties, thus improving their interpretation and understanding. Such an approach increases the quality and accuracy of the applied metadata and is essential for the effective search and image retrieval[5]. The use of structured vocabularies harmonizes the presentation of cultural heritages and two actual fields of research and application of today's Web: application of concepts of the Semantic Web and search of images.The concepts of the Semantic Web can help:
- search engines to understand and accept contents of the Web and
- Web users to createmore precise search queries in an easier way

The responses to the refined queries contain less unnecessary and useless information that are offered to Web users. To achieve this, it was necessary:
- to annotate semantic meanings to an image, by linking to existing semantic resources on the Web and
- to expose the annotated images to search engines, but in a preferable manner.

There are more standards that regulate the space of the Semantic Web. Also, more technologies and software tools support the implementation of the Semantic Web. Thus, the choice is not unique. In this article we present an applicable combination of standards together with the corresponding technologies.

3 Representing images as a part of web of data

The main motive for publishing some content onthe Web is to spread out information, ideas and attitudes to a wider or narrower group of Web users. The understanding of new, semantically annotated documents is significantly improved, if we connect them with the already existing and meaningful data on the Web. In this way, the process of linking Web data transforms the initial Web into the Web of Data. Linked data create a graph that helps machines to find, select, follow and index Web contents[6]. The same is true in the case of images. So, it is recommended to enrich the visual content of an image by semantic meaning in the way that is understandable to both people and machines. Further, the use of Semantic Web technologies for publishing semantically enriched images of outsider artworks will be explained.

3.1 RDF, Linked Data and JSON-LD. The Resource Description Framework (RDF) and Linked Data (LD) get the basic functionalities of the Web of Data. RDF defines Web resources by means of triples: subject, predicate, object. The phrase "Web resource" assigns a Web content that has the unique Internet address, defined by Internationalized Resource Identifiers (IRI)[7].The use of RDF providesthe creation of LD. Usually there are links between different LD. When they exist, links are established using the IRIs and the HTTP protocol[6]. In other words, the RDF in the Semantic Web takes over the role of HTML documents in the initial Web[8]. It means that the Semantic Web is the congregation of connected data, instead of linked HTML documents. So, if annotated images, or some other Web contents use LD resources, they are simultaneously integrated into the Web of Data.The release of the LD from the proprietary licenses increases the distribution and acceptance of the LD,thus they become available for use as the Linked Open Data (LOD). The ownership release stimulates interconnecting sets of LODs and the constitution of LOD network[1].

Cultural institutions such as Galleries, Libraries, Archives and Museums (GLAM) use benefits of the existing cultural LOD, adopting description and presentation of their cultural properties to the Web of Data. Figure 1 presents an example of cultural data as a part of LOD.

In the Project "Marginal Art in Serbia" we explored Web promotion of outsider art to support its transition from the artistic margins into the mainstream. To get advantages of the Web of Data, we linked the dataon marginal artworks to the cultural LOD.

Figure 2 illustrates a part of the RDF graph, based on semantically annotated data for an image from our project.
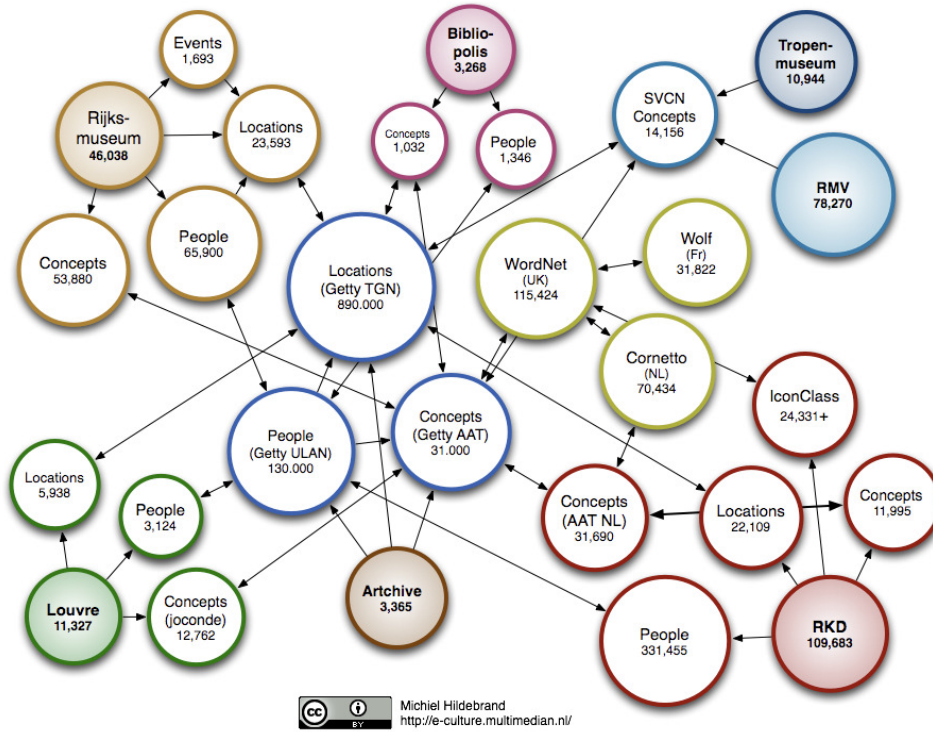
---

[1] http://lod-cloud.net/

Figure 1: "Culture Datacloud" by Michiel Hildebrand
(https://commons.wikimedia.org/wiki/File:Culture_Datacloud.png#/media/File:Culture_Datacloud.png)
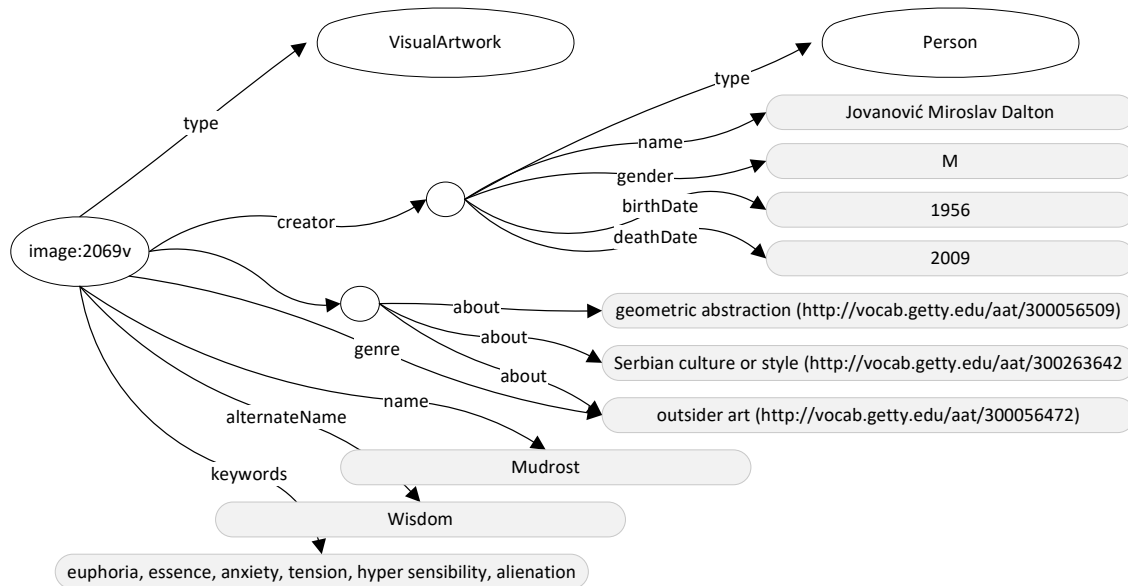

Figure 2 Example of the part of RDF triples that describes semantically annotated image 2069v.jpg from the project "Outsider Art in Serbia". The links point to other data sources on the Web, including Getty vocabularies.

Publishing abstract RDF models into any acceptable Web format requires serialization. Some of popular serialization formats are: JSON-LD[2], RDF/XML, N-Triple, N-Quads and Turtle. We prefer to use JSON-LD[3], as we agree that the JSON-LD format is more appropriate for creating and reading Linked Data [9].Below is the part of the JSON-LD source code that corresponds to the serialized RDF model, presented by the Figure 2.

```
{
  "@context": "http://schema.org/",
  "@type": "VisualArtwork",
  "name": "Wisdom"
  "alternateName": "Mudrost",
  "artMedium": "mixed media (http://vocab.getty.edu/aat/300163347)",
  "artform": "painting (http://vocab.getty.edu/aat/300054216)",
  "artworkSurface": "canvas (http://vocab.getty.edu/aat/300014078)",
  "genre": "outsider art (http://vocab.getty.edu/aat/300056472)",
  "image": "http://www.marginalart.rs/illustrations/2069v.jpg",
  "keywords": "euphoria, essence, anxiety, tension, hyper sensibility, alienation",
  "about": [
   "geometric abstraction (http://vocab.getty.edu/aat/300056509)",
   "outsider art (http://vocab.getty.edu/aat/300056472)",
   "Serbian culture or style (http://vocab.getty.edu/aat/300263642"
  ],
  "creator": [
        { "@type": "Person",
        "name": "Jovanović Miroslav Dalton",
        "birthDate": "1956",
        "deathDate": "2009",
        "gender": "M" }
   ]
}
```

In this example JSON-LD refers to Schema.org vocabulary by IRI identifier. Also, it includes Schema.org properties to connect to LOD.

In the next section the main properties of LOD vocabularies will be presented and the importance and the role of the vocabularies within the Semantic Web will be highlighted.

3.2 Semantic Image Annotation Using LOD Vocabularies and Schema.org. Vocabularies systematize words and phrases, in forms of thesauri, taxonomies or ontologies. Some of them are alphabetically organized lists of words. Others are subject oriented and hierarchically organized. More complex vocabularies include networked concepts. [10][11].A significant role of vocabularies in the Semantic Web is to improve the understanding of Web contents. Vocabularies help to categorize Web documents and support the creation of structured data, thusimproving the precision of Web searching.

One of the main objectives of our Project "Marginal Art in Serbia" was to semantically annotate images of marginal artworks. Following recommendations and benefits of vocabularies [12][8], we decided to use the already accepted and authoritative LOD vocabularies. We based our decision on the facts that:
-   The use of existing vocabularies increase interoperability and
-   Each concept from the vocabulary becomesa data integration node.
We used LOD Getty vocabularies to interpret marginal artworks, such as:

- Art and Architecture Thesaurus (AAT)[4] - as it contains the concepts of art, architecture, decorative art, archival material, etc.
- Getty Thesaurus of Geographic Names (TGN)[5] - as it contains geographical names and locations.

Getty vocabularies are in compliance with the standards in the domain of art and they offer the localization. [11]. Figure 3 shows the position of the concept "outsider art" within the hierarchical tree view of the AAT vocabulary.
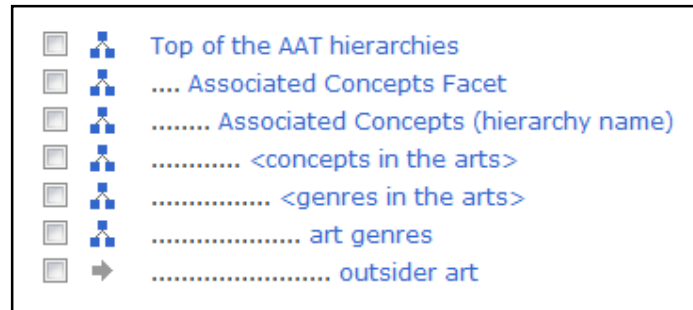

Figure 3: The position of the concept "outsider art" in the AAT Vocabulary[6]

To present and interpret artworks of marginal artists to specialists and general Internet users, we connected our semantically annotated images of marginal art to the LOD. In the final step we used the Schema.org vocabulary. The Schema.org vocabulary allowed us to express the selected metatags from the used LOD in a form that is acceptable and desirable by search engines.

Search engines, such Google, Yahoo!, Bing and Yandex, initiated the creation of Schema.org[7] vocabulary. The use of Schema.org vocabulary, together with structured data markup: Microdata, Microformats, RDFa and JSON-LD, improve determination and interpretation of Web contents, obtaining more accurate searching results and better positioning in search engine indexes. Besides, Schema.org is a Linked Open Vocabulary. Therefore, it is reasonable to expect that it could be one of dominant nodes within the LOD, and it is advisable to use it for the purpose of Web content annotation [13].

While developing the Project "Marginal Art in Serbia", we combined the described technologies and formats for semantic image enrichment in such a way to make it recognizable to people and machines. Figure 4 presents the screenshot of the HTML page that presents the image from project.
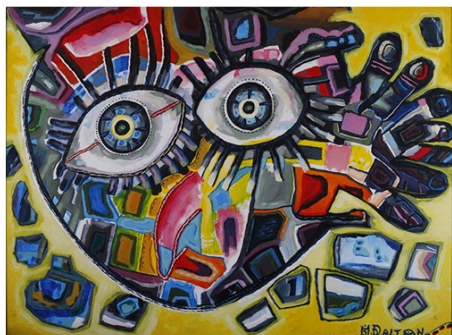
---

[4] http://www.getty.edu/research/tools/vocabularies/aat/index.html

[5] http://www.getty.edu/research/tools/vocabularies/tgn/index.html

[6] http://vocab.getty.edu/aat/300056472

[7] http://schema.org/

Figure 4: The screenshot of the HTML page that presents the image from project
(http://www.marginalart.rs/outsider-art-image/2069v)

4 Semantic annotation of outsider art images

The evaluation of artworksis not usually exact. The acceptance of values of an artwork depends on the moment of creation, socio-political surroundings, artist's status, actual artistic trends, etc. The same is true for outsider artworks. Outsider artists have no formal artistic education and are not exposed to direct influence of the school and artistic rules. For that reason a synonym of the "outsider art" is the "art brut". The creativity of artists has been on the margins of art for a long time, hence the term "outsider art" has another synonym: "marginal art". At the end of the twentieth century, increasing interest in the works of outsider artists began. Itinfluenced the transfer of outsider art from artistic margins to the current mainstreams.

MNMA has a significant collection of outsider artworks whichare subject to protection, studying and digitization. Outsider artworks contain layered messages. Even at the first glance, the visual layer suggests messages that are psychological reflections of the artist at the momentof creation. Thus, the presentation and interpretation of an outsider artwork is not a simple task. For that reason MNMA started the Project "Outsider Art in Serbia". The main objectives of the Project were interpretation, Web publishing and the promotion ofoutsider artworks from the Outsider MNMA collection.The Project included:

- recognition and interpretation of outsider artwork messages,
- semantic annotation of outsider artwork images and
- Web promotion of the most important artworks from the Outsider collection of MNMA[14].

The multidisciplinary project team includedart historians, psychologists and IT specialists. The task of art historians was to recognize artistic and aesthetic values of artworks. Since outsider artworks express intensive psychological conditions of the artist[15], psychologists recognized messagesby analysing visual contents, such as fear, joy, irony, flight, fight, etc. In the previous sections of this article, we presented the benefits and importance of linking semantically annotated images to the LOD vocabularies and Schema.org. Thus, we had to identify attributes that describe an

artwork by using the appropriate metatags from the selected vocabularies. We provided a set of information about the artwork, author and links within the Web environment:
- Artwork (image): title, date of creation; art form; description; keywords; dimensions; medium; surface; genre; Web address of the image; license note.
- Artist: name; gender; date, place and country of birth and of death.
- Museum: name; address and geo-coordinates, Web URL;
- Links to Wikipedia pages about the Museum and artist, if they exist.

The tasks of IT specialists were to create software solutions that implemented the selected metatags of an image, connecting them to the specified LOD vocabularies and other semantic contents on the Web. To do this, we applied the technologies and dictionaries presented in the Chapter 3. Finally, semantically annotated images of outsider artworks from the collection of MNMA were published on the Web site http://www.marginalart.rs. Figure 5 presents the overview of our approach for semantic enriching images of outsider artworks.
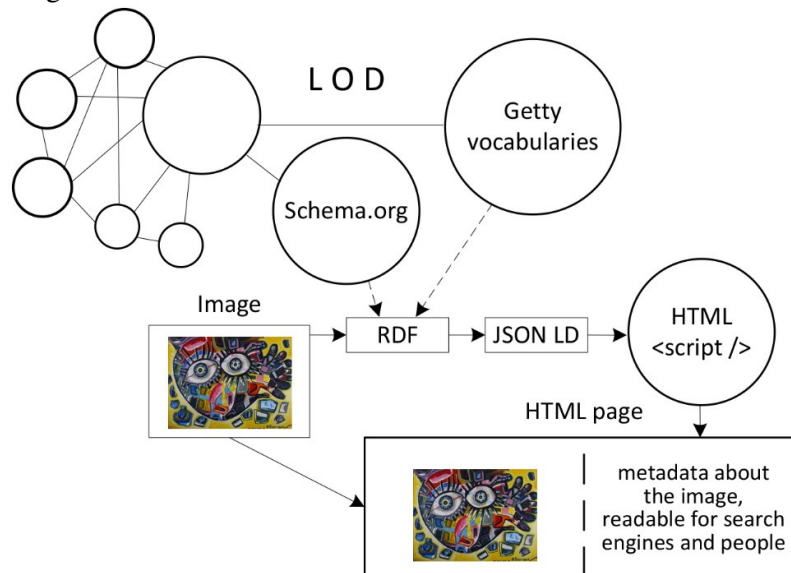

Figure 5: The overview of the approach for semantic enriching images of outsider artworks from the project "Outsider Art in Serbia"

The last phase of the project explores the effects of the applied semantic annotation of the outsider images within the Web environment.

First, Google has recognized and processed the structured schemes of our semantically annotated images based on Schema.org vocabulary. Figure 6 presents a screenshot of Google Webmaster Tools that illustrates how Google interprets the semantic data for an image from the Project "Outsider art in Serbia".

Figure 6: Example of the recognized semantic data at the Google Search Console

Second, the suggestion that Google favours structured data seems to be true. The Images of artworks from the Collection of Outsider MNMA take high ranking in the Google indexes within the category of marginal art. This is obvious even in the case of extremely competitive searching terms, such as fear, sadness, happiness, alienation, irony, etc. Figure 7 presents the first Google resulting page of images that correspond to the search phrase: "marginal art fear".
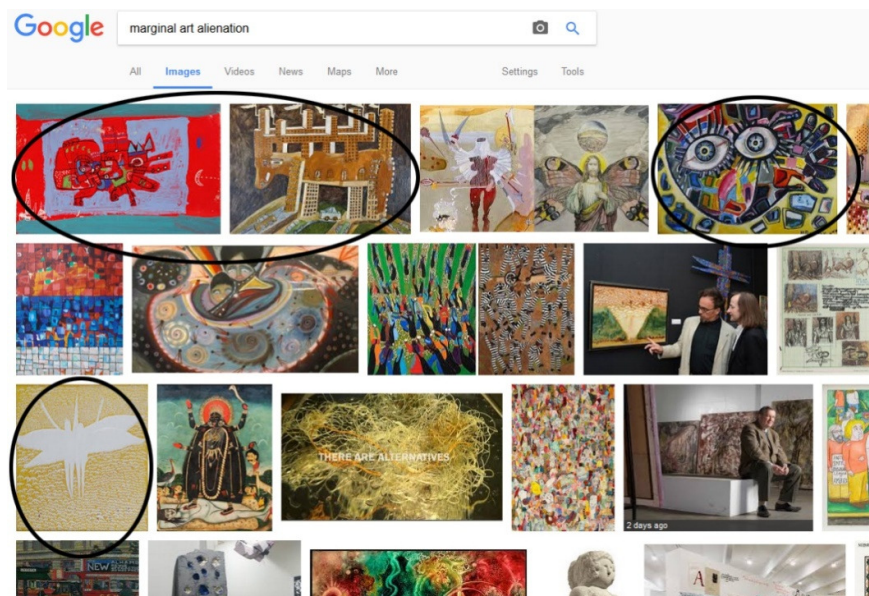

Figure 7: Google results of searching images for the phrase: "marginal art alienation", 12 March 2017, our images are marked by ellipses.

Conducting the Project "Outsider Art in Serbia" we canconclude that the described procedure, illustrated by Figure 5, support all phases of the project. The Use of LOD vocabularies direct the selectionof words and phrases to describe the layered artistic messages in accordance with the previously verified professional terms. Also,

LOD vocabularies classify the annotated outsider artworks into the corresponding categories and thematic contexts.

5 Conclusion and future works

It seems that publishing images on the Web is a simple task. Thereforethe Web has become an extremely large repository of thematically heterogeneous images. At the same time, today's technical capabilities spread the growing digitization trend of cultural artefacts and publishing of digitized images of cultural heritage on the Web. The realization of the GLAM project confirms such a practice.

In this article we discuss the technologies and tools that are suitable for semantic annotation images of cultural artefacts. We applied the conclusions of our research in the Project "Outsider Art in Serbia". Our approachis illustrated in Figure 5.

We used Getty LOD and Schema.org vocabularies to annotate the images of outsider artworks. Google recognized semantically enriched visual contents of images as structured data, assigning them high rankings in its indexes.

Based on our experience, we conclude that it is advisable to use LOD vocabularies. Our research and application show that the complex image data, enriched by using LOD and Schema.org vocabularies and serialized by JSON-LD embedded into the HTML pages (Figure 4), became visible and understandable to both machines and people.

The findings and results achieved within the described Project, motivate us to proceed with further activities of semantic image enrichment to some other categories of digitalized heritage. Regarding the diversity of possible applications, we plan to focus on the application of the described method to highlight mutual relationships between some kinds of tangible and intangible cultural heritage.

References

[1]   J. Jokilehto, "DEFINITION OF CULTURAL HERITAGE REFERENCES TO DOCUMENTS IN HISTORY," 2005.

[2]   "Linked Open Vocabularies." [Online]. Available: http://lov.okfn.org/dataset/lov. [Accessed: 11-Mar-2017].

[3]   M. Aćimović and S. Negovanović, *Preporuke za stvaranje i upravljanje digitalnom foto-dokumentacijom u institucijama zaštite kulturnog nasleđa Srbije*. Beograd: Službeni glasnik, 2015.

[4]   J. Itten, *The Art of Color: The Subjective Experience and Objective Rationale of Color*. John Wiley & Son, 1997.

[5]   B. Schandl, B. Haslhofer, T. Bürger, A. Langegger, and W. Halb, *Linked Data and multimedia: the state of affair*s, Multimed. Tools Appl*., 59: 2(2012), 523–556

[6]   C. Bizer, T. Heath, and T. Berners-Lee, *Linked data-the story so far*, Int. J. Semant. Web Inf. Sys*t.*, 5:3(2009), 1–22

[7]   L. Sauermann, R. Cyganiak, D. Ayers, and M. Völkel, *Cool URIs for the Semantic Web*, 2008. [Online]. Available: https://www.w3.org/TR/cooluris/. [Accessed: 29-Oct-2016].

[8]   T. Heath and C. Bizer, *Linked Data - Evolving the Web into Global Data Space*, Synthesis Lectures on the Semantic Web: Theory and Technology, J. Hendler and F. van Harmelen, Eds. Morgan & Claypool, 5(2011), 1–136.

[9]   M. Sporny, D. Longley, G. Kellogg, M. Lanthaler, and N. Lindström, *JSON-LD 1.0 A JSON-based Serialization for Linked Data*, W3C Recommendation, 2014.

[Online]. Available: https://www.w3.org/TR/json-ld/.

[10] M. L. Garshol, *Metadata? Thesauri? Taxonomies? Topic Maps!,* J. Inf. Sci., 30:4(2004), 378–391

[11] P. Harpring, *Introduction to controlled vocabularies: terminology for art, architecture, and other cultural works*. Getty Research Institute, 2010.

[12] B. Hyland, G. Atemezing, and B. Villazon-Terrazas, *Best Practices for Publishing Linked Data*, W3C Working Group Note, no. January 2014, 1–22

[13] P. Mika, *On Schema.org and Why It Matters for the Web*, IEEE Internet Comput., 19:4(2015), 52–55

[14] N. Krstic, *Outsider Art in Serbia*. Museum of Naïve and Marginal Art, 2014.

[15] N. Krstic, *Art in Spiritual Exile*. Museum of Naïve and Marginal Art, 2015.